

# **Dynamic Effects on the Stability of International Environmental Agreements**

**Aart de Zeeuw<sup>1</sup>**

Department of Economics and CentER, Tilburg University

The Beijer Institute of Ecological Economics

P.O. Box 90153, 5000 LE Tilburg, the Netherlands

phone: +31-13-4662065; fax: +31-13-4663042

e-mail: [A.J.deZeeuw@uvt.nl](mailto:A.J.deZeeuw@uvt.nl)

## Abstract

In terms of the number of signatories, one observes both large and small international environmental agreements. The theoretical literature, based on game theory, discusses different concepts and mechanisms regarding stability of coalitions. The conclusion has been reached that in all circumstances, under farsightedness, both large and small stable coalitions can occur. This theory is based on behavioural reaction patterns but does not take account of the interaction with the dynamics of emission adjustments. This paper shows that when these two dynamical processes are integrated, large and small stable coalitions can still occur but only if the costs of emissions are relatively unimportant as compared to the costs of abatement.

JEL codes: Q2, C70, F42

Key words: IEA's, coalitional stability, dynamics

## 1. Introduction

Global environmental problems such as ozone depletion and climate change require voluntary cooperation by sovereign states to internalise the negative externalities of cross-border emissions. In the last decades, so-called International Environmental Agreements (IEA's) were signed, but the number of signatories varies considerably (e.g., [11]). The Montreal Protocol (1987) with the purpose to phase out CFC's (that cause depletion of the ozone layer) has been signed and ratified by 181 countries. However, the Kyoto Protocol (1997) with the purpose to reduce greenhouse gas emissions (that cause climate change) is a different story. The number of countries that will reduce emissions in the first phase is much smaller, and especially after the withdrawal of the USA in 2001, the basis of this agreement is very weak. It came into force only recently by the ratification of Russia.

The theoretical literature on this issue has mainly focused on stability concepts for coalitions (see [11] for an overview). The dominant strand in this literature is based on the idea of internal and external stability, meaning that an individual country neither has an incentive to leave nor to join the coalition ([13,5,2]). For most models in this tradition, the outcome is rather grim: the size of the stable coalition is very small. A more optimistic view on the size of stable coalitions results if some trigger mechanism is introduced that can change the position or the behaviour of the other countries. If deviation of one country would induce the coalition to fall apart, the mere threat of loosing all cooperative benefits may deter deviations in the first place and support cooperation. This idea is at the heart of the trigger mechanism in repeated games ([12]). It is also the main idea for the  $\gamma$ -core concept that is rooted in

cooperative game theory and that is used to investigate stability of the grand coalition ([6]).

These two strands of literature take opposite assumptions on what happens in case a country leaves the coalition. Either it is assumed that the rest of the coalition remains intact or it is assumed that the whole coalition breaks down. Farsightedness (e.g., [16]) is a concept that advocates behavioural assumptions that are in a sense midway between these two models. The idea is that if a country leaves, it may trigger other countries to leave as well until a new stable situation is reached. In this case a country compares its initial position with its position at the end of this process. It can be shown that farsightedness leads to a set of large and small stable coalitions ([8,10]). In a way the two approaches, described earlier, are reconciled by this concept. Moreover, an important conclusion is that countries can coordinate on a large coalition that is stable.

Dynamic aspects are mostly ignored in this literature. Behavioural reaction patterns are dynamic, of course, but abatement processes are usually dynamic as well. In most of the models it is assumed that countries reduce emissions in one step, but this is not realistic and also not rational, as can be seen from a standard optimal control model with convex costs where the target is reached in steps. In practice countries take time to adjust their emission levels. For example, in the Kyoto Protocol the 15 countries of the European Union plan to reduce greenhouse gas emissions with 8 % by 2008-2012, as compared to 1990 levels. This paper investigates how the interaction between the behavioural reaction pattern and this adjustment process affects the results on the stability of coalitions.

The concept of farsightedness is sufficiently rich to allow for a set of large and small stable coalitions. Suppose that the countries coordinate on the largest stable coalition. Deviation is deterred because under farsightedness it is not in the interest of a country to become an outsider to a smaller stable coalition. If welfare is discounted over time, this statement is still true, provided the discount factor is high enough. However, if the other members of the coalition make adjustments before a deviation is detected, the game changes over time and then it is not so clear that deviations will be deterred.

This paper uses an optimal control model with the objective to phase out emissions in the long run at the lowest costs. At each time step the countries choose some level of abatement, like investment in green technology, in order to lower the emission levels. In the absence of cooperation, this adjustment process is slower and has higher costs than in case (some) countries cooperate. It will be shown that a one-shot game of this type always has large stable coalitions under farsightedness. However, in the dynamic game it depends on the relative costs of abatement and emission levels whether large coalitions can be sustained or not. This result may give more insight into why certain international environmental agreements are successful, in terms of membership, and others are not.

## **2. The model**

The basic model is a simple abatement model. In the initial state of the world, global emissions are at a level  $e_0$ . Under business-as-usual, without environmental concern, global emissions will grow over time with a factor  $b$ .<sup>2</sup> There are  $n$  countries ( $n > 2$ ).

At each point in time, each country can choose an abatement level  $a_i$ ,  $i = 1, 2, \dots, n$ , like an investment in green technology or a structural change. This yields the following dynamics for the level of emissions

$$(1) \quad e(t+1) = (1+b)e(t) - \sum_{i=1}^n a_i(t), t = 0, 1, \dots, e(0) = e_0.$$

The emission level is damaging but it is also costly to abate. A simple cost indicator is given by

$$(2) \quad C_i = \sum_{t=0}^{\infty} \delta^t [ \frac{1}{2} a_i^2(t) + \frac{1}{2} p e^2(t) ], i = 1, 2, \dots, n, p > 0,$$

where  $p$  denotes the relative weight attached to the damage costs as compared to the abatement costs and  $\delta$  denotes the discount factor.

This is an optimal control model where the level of emissions is driven down to zero, regardless of whether the countries cooperate or not. In that sense it differs from the typical international pollution control model with a stock pollutant (e.g., [15]) where the cooperative steady state is not the same as the non-cooperative steady state. In the case of climate change, it would be more appropriate to formulate a stock pollutant model, because damage is related to the stock of greenhouse gases. The message of the paper would not change, however, but the analysis is much more complicated. Moreover, the Kyoto Protocol is an agreement on the level of emissions, and not on the level of the stock of greenhouse gases. Note that the model still has a “state”. The state of the system is the information that is needed to determine future optimal or equilibrium strategies. In this model the level of emissions is the state. In a stock pollutant model the state would be the stock of greenhouse gases. The model in this paper describes a situation where the target in the long run is to reduce emissions to zero but under cooperation this target is reached faster and with lower costs than in

the absence of cooperation. This will become clear in section 4 where the model is solved. Furthermore, note that in order to lower the level of emissions, first of all abatement has to compensate the business-as-usual growth in emissions. This implies that the value of the growth factor  $b$  will affect the levels of abatement, but it will not affect the qualitative results of the model. Finally, the  $n$  countries are assumed to be identical. This is not very realistic, of course, especially in the situation of climate change, but in this way we can abstract from issues like transfers between countries within a coalition or a possible relationship between size and membership. We want to focus on the number of countries that join the coalition. An extensive literature exists (e.g., [4,17,16]) on general coalitional structures but we restrict the analysis to one coalition with all other countries as individual outsiders. This corresponds to the current practice of international environmental agreements, although recently competing agreements with a different focus are developing such as the technology agreement, led by the USA, as an alternative for the Kyoto Protocol.

### 3. Stability concepts

In order to be able to discuss the different stability concepts in a simple framework, we start with a one-shot static version of the model with emission level  $e$ ,  $b = 0$ , and cost indicators

$$(3) \quad C_i = \frac{1}{2}a_i^2 + \frac{1}{2}p(e - \sum_{j=1}^n a_j)^2, i = 1, 2, \dots, n.$$

Suppose that  $m$  ( $< n$ ) countries form the coalition. This implies that these  $m$  countries (called members) jointly minimize the sum of their costs, whereas the other countries (called outsiders) just minimize their own costs. It follows that first-order conditions for the equilibrium levels of abatement for members  $i$  and outsiders  $j$  are given by

$$(4) \quad a_i - mp(e - \sum_{k=1}^n a_k) = 0, i = 1, \dots, m,$$

$$(5) \quad a_j - p(e - \sum_{k=1}^n a_k) = 0, j = m + 1, \dots, n.$$

Adding up and some simple manipulations yield the resulting level of emissions and the equilibrium abatement levels:

$$(6) \quad e - \sum_{k=1}^n a_k = \frac{1}{1 + (m^2 + n - m)p} e,$$

$$(7) \quad a_i = \frac{mp}{1 + (m^2 + n - m)p} e, i = 1, \dots, m,$$

$$(8) \quad a_j = \frac{p}{1 + (m^2 + n - m)p} e, j = m + 1, \dots, n.$$

It follows that the costs in equilibrium for a coalition member ( $C^c$ ) and for an outsider ( $C^o$ ) are given by

$$(9) \quad C^c = \frac{1}{2} \frac{p + m^2 p^2}{(1 + (m^2 + n - m)p)^2} e^2,$$

$$(10) \quad C^o = \frac{1}{2} \frac{p + p^2}{(1 + (m^2 + n - m)p)^2} e^2.$$

A few things are immediately clear: for  $m = 0$  or  $m = 1$  one gets the standard Nash equilibrium between individual countries and for  $m = n$  one gets the full-cooperative outcome, with higher abatement levels and lower total costs. Note that for any size of the coalition  $m > 1$  the costs of the outsiders are lower than the costs of the coalition members and also lower than the costs in the Nash equilibrium (free-rider benefits).

The usual approach to the theory on stable International Environmental Agreements (e.g., [5,2]) is based on the ideas developed for cartel stability ([1]) and requires what

is called *internal* and *external* stability. Internal stability means that a country does not have an incentive to leave the coalition, that is the costs of an outsider to the coalition of size  $m-1$  must be larger than (or equal to) the costs of a member of the coalition of size  $m$ . External stability means that a country does not have an incentive to join the coalition, that is the costs of a member of the coalition of size  $m+1$  must be larger than the costs of an outsider to the coalition of size  $m$ . This corresponds to the Nash equilibrium of the so-called open-membership game in which the countries first decide on whether they want to be a member of the coalition or not and then choose their abatement levels ([4,17,11]). It is straightforward to derive from (9) and (10) that

$$(11) \quad C^c(m) > C^o(m-1)$$

for all  $p$ ,  $n$ , and  $m > 2$ , so that all coalitions larger than 2 are not internally stable. The same inequality (for  $m = 3$ ) also implies that a coalition of size 2 is externally stable. It follows that the size of the coalition that is both internally and externally stable is either 1 or 2. Straightforward calculations show that if  $p \leq 1/(n-4+2\sqrt{n^2-3n+3})$  the size is 2 and otherwise the size is 1.

The general conclusion of this analysis is that the size of the stable coalition is small. In this model the size is not larger than 2. Moreover, the size is only 1 if the relative weight  $p$  of the damage costs as compared to the abatement costs is high. As can be seen from (4) and (5), the parameter  $p$  is also an indicator of the steepness of the best-reply function. Hoel [13] and Carraro and Siniscalco [5] note that the slope of the best-reply function indicates by how much the outsiders offset the extra abatement of the coalition members (leakage effect). Based on this it is argued that the steeper the slope, the larger the incentive to deviate from the coalition. Finus [11] extends this to a broader set of models and concludes: the steeper the best-reply function, the lower



the number of participants. Barrett [2] finds the opposite result but with a Stackelberg model in which the coalition moves first and can take the reaction of the outsiders into account. Moreover, the Stackelberg model allows for a wider range of outcomes, not only 1 or 2. Barrett [2] uses this result to give an explanation for the difference in participation rate between the Montreal Protocol and the Kyoto Protocol.

The conclusion that the size of the stable coalition is very small was challenged from different angles. Chander and Tulkens [6] use the  $\gamma$ -core concept from cooperative game theory to argue that the grand coalition can be stable. It is assumed that when a sub-coalition of size  $m$  deviates the initial coalition falls apart and a game is played between the sub-coalition of size  $m$  and the other countries individually. In the symmetric case this is precisely the game described above with costs  $C^c(m)$  and  $C^o(m)$ , respectively. Because  $C^c(m) > C^c(n)$  for all  $m < n$ , sub-coalitions do not have incentives to deviate and the grand coalition is stable in that sense. This mechanism is similar to trigger strategies in repeated games ([12]). In that model it is assumed that individual deviations are punished by a switch to non-cooperative behaviour. The threat of loosing the cooperative benefits in the future is sufficient to deter deviations and to support the grand coalition, provided the discount factor is high enough. These models are criticized because the assumptions are strong, but weaker assumptions also prove to support large stable coalitions.

Recently the concept of *farsightedness* was introduced (e.g., [7,16]) and applied to International Environmental Agreements ([8,10]). The basic idea is that countries realize that a deviation may trigger further deviations but they do not necessarily assume a full breakdown. In that perspective, the mechanism of internal stability is

too weak, since this assumes that no further deviations take place, but the mechanism of the  $\gamma$ -core is too strong, since this assumes that the initial coalition falls apart completely. The idea is furthermore that such a sequence of deviations comes to an end when a (new) stable situation is reached. If an outsider in that new situation is worse off than a coalition member in the initial situation, deviations are deterred.

It is useful for the sequel to label these different stability concepts.<sup>3</sup> The requirement of internal and external stability will be called *myopic stability* and the requirement of farsightedness will be called *static farsighted stability*. For the game in this section it was found that for  $p$  sufficiently small, only the size 2 coalition is myopically stable. The coalition of size 2 therefore satisfies static farsighted stability, if it is assumed that the coalition of size 1 belongs to the set of static farsighted stable coalitions.<sup>4</sup> However, coalitions larger than 2 may be elements of this set as well. The coalition of size 3 is not myopically stable and therefore also not stable in the sense of static farsightedness, but the coalition of size 4 may be stable in that sense. It may be better to be a member of that coalition than an outsider to the coalition of size 2. It is not difficult to show that this is indeed the case. If the total number of countries  $n$  is large enough, this process continues and usually generates a bigger set of stable coalitions. The largest one is usually close in number to the grand coalition. Further details will not be investigated here but will be left for the analysis of the dynamic model in the next section.

The general conclusion from this section is that reasonable behavioural assumptions lead to large stable coalitions. The usually grim picture that (in a Nash equilibrium context) the size of the stable coalition is only 1 or 2 has disappeared. However, in the

next section it will be shown that this conclusion does not generally hold in a dynamic model in which detection of deviations takes time. This is unfortunate but also has an advantage: the size of the stable coalition depends again on the parameter  $p$ , as in the model with myopic stability, but now in a much richer framework that takes account of farsightedness and dynamic emission adjustments. The results can be connected to the characteristics of the environmental problem and thus used to understand the size of International Environmental Agreements in practice.

#### **4. Stability in a dynamic context**

Let us return to the model (1) - (2) presented in section 2. This is called a difference game because the state transition is given by the difference equation (1). It is assumed that the countries can observe the emission levels and can therefore condition their abatement activities  $a_i$  at time  $t$  on the level of emissions  $e$  at time  $t$ . Furthermore, it is assumed that the countries cannot credibly commit to a strategy beforehand and may adjust their strategy at each point in time. This implies that we look for the feedback Nash or Markov perfect equilibrium. The parameters in the state transition (1) and in the cost indicators (2) are time-independent. The solution of this difference game will therefore be stationary, meaning that the equilibrium abatement activities will only be a function of the level of emissions and not a function of time. The levels will change over time, of course, but the functional relationship between abatement and emissions remains the same. The feedback Nash or Markov perfect equilibrium can be found by solving the Hamilton-Jacobi-Bellman or dynamic programming equations in the value functions of the countries (e.g., [9]).

Suppose again that  $m$  countries form a coalition and that a Nash equilibrium results between the coalition and the other countries individually. For the time being,  $m$  is fixed in order to derive the value functions as a function of  $m$ . These value functions will subsequently be used to analyse possible changes in the coalition size  $m$ , as a consequence of deviations. The dynamic programming equations are

$$(12) \quad \frac{1}{2}mk^c e^2 = \min \left[ \sum_{i=1}^m \frac{1}{2}a_i^2 + \frac{1}{2}mpe^2 + \frac{1}{2}\delta mk^c ((1+b)e - \sum_{k=1}^n a_k)^2 \right],$$

$$(13) \quad \frac{1}{2}k^o e^2 = \min \left[ \frac{1}{2}a_j^2 + \frac{1}{2}pe^2 + \frac{1}{2}\delta k^o ((1+b)e - \sum_{k=1}^n a_k)^2 \right], j = m+1, \dots, n$$

where  $V^c(e) = \frac{1}{2}k^c e^2$  and  $V^o(e) = \frac{1}{2}k^o e^2$  denote the value functions for a member of the coalition and for an outsider, respectively. It follows that the first-order conditions for the equilibrium levels of abatement are

$$(14) \quad a_i - \delta mk^c ((1+b)e - \sum_{k=1}^n a_k) = 0, i = 1, \dots, m,$$

$$(15) \quad a_j - \delta k^o ((1+b)e - \sum_{k=1}^n a_k) = 0, j = m+1, \dots, n.$$

Adding up and simple manipulations lead to the next-period level of emissions and to the feedback Nash equilibrium abatement strategies:

$$(16) \quad (1+b)e - \sum_{k=1}^n a_k = \frac{1}{1 + \delta(m^2 k^c + (n-m)k^o)} (1+b)e,$$

$$(17) \quad a_i = \frac{\delta mk^c}{1 + \delta(m^2 k^c + (n-m)k^o)} (1+b)e, i = 1, \dots, m,$$

$$(18) \quad a_j = \frac{\delta k^o}{1 + \delta(m^2 k^c + (n-m)k^o)} (1+b)e, j = m+1, \dots, n.$$

Substitution in the dynamic programming equations (12) - (13) yields the following set of two equations for the (positive) parameters  $k^c$  and  $k^o$  of the two value functions:

$$(19) \quad k^c = p + \frac{\delta k^c (1 + \delta m^2 k^c)}{(1 + \delta(m^2 k^c + (n - m)k^o))^2} (1 + b)^2,$$

$$(20) \quad k^o = p + \frac{\delta k^o (1 + \delta k^o)}{(1 + \delta(m^2 k^c + (n - m)k^o))^2} (1 + b)^2.$$

As in the static case in section 3, for  $m = 0$  in equation (20) one gets the equation for the parameter  $k^{na}$  of the value function for the feedback Nash equilibrium between individual countries, namely

$$(21) \quad k^{na} = p + \frac{\delta k^{na} (1 + \delta k^{na})}{(1 + \delta n k^{na})^2} (1 + b)^2,$$

with abatement strategies

$$(22) \quad a_i^{na} = \frac{\delta k^{na}}{1 + \delta n k^{na}} (1 + b) e, i = 1, \dots, n,$$

and for  $m = n$  in equation (19) one gets the equation for the parameter  $k^{fc}$  of the value function for the full-cooperative solution, namely

$$(23) \quad k^{fc} = p + \frac{\delta k^{fc}}{1 + \delta n^2 k^{fc}} (1 + b)^2,$$

with abatement strategies

$$(24) \quad a_i^{fc} = \frac{\delta n k^{fc}}{1 + \delta n^2 k^{fc}} (1 + b) e, i = 1, \dots, n.$$

It is not difficult to show (without explicitly solving equations (21) and (23)) that

$$(25) \quad k^{fc} < k^{na} < n k^{fc}.$$

The first inequality in (25) implies that for any initial global emission level, total costs are lower under full cooperation than in the feedback Nash equilibrium between individual countries. The second inequality in (25), with equations (22) and (24),

implies that abatement levels are always higher under full cooperation. This is to be expected and confirms the claims made when introducing the model in section 2. The situation for a coalition of size  $m$  with individual outsiders, however, is much more complicated.

The structure of the model is relatively simple. The value functions have a quadratic form without linear or constant terms, so that only one parameter per value function has to be determined. Nevertheless, the set of equations (19) - (20) is still complicated and cannot be solved analytically. We have to resort to a numerical solution, but first a precise definition of what we mean by stability in a dynamic context is needed.

In principle, the value functions  $V^c(e_0) = \frac{1}{2}k^c e_0^2$  and  $V^o(e_0) = \frac{1}{2}k^o e_0^2$  can be analysed in the same way as the cost functions  $C^c$  and  $C^o$ , given by equations (9) - (10). Again, it can be shown that the requirement of myopic stability is only satisfied for very small coalitions but static farsighted stability yields a set of large and small stable coalitions. This approach, however, does not take account of the dynamics of the problem. More specifically, it takes time to detect deviations. Therefore, a deviator enjoys extra free-rider benefits before the deviation is detected and this may change the situation. In a repeated game the stability properties would not change, if the discount factor is high enough. For example, suppose that the static game of section 3 is repeated and that the coalitions of size  $m$  and  $m^+$  ( $m > m^+$ ) are subsequent elements in the set of coalitions that satisfy static farsighted stability. It follows that  $C^c(m) < C^o(m^+)$ . Furthermore, suppose that a deviator has costs  $C^d(m) < C^c(m)$  before detection and becomes an outsider to the coalition of size  $m^+$  afterwards. Deviation is deterred if

$$(26) \quad \frac{1}{1-\delta} C^c(m) < C^d(m) + \frac{\delta}{1-\delta} C^o(m^+)$$

and this holds if the discount factor is high enough, i.e. if

$$(27) \quad \delta > \frac{C^c(m) - C^d(m)}{C^o(m^+) - C^d(m)}.$$

However, the game introduced in section 2, and analysed in this section, changes over time. Due to the abatement activities of the countries in the first period, the level of emissions and therefore the costs decrease. The future values of coalition members and outsiders do not only depend on the parameters  $k^c$  and  $k^o$  of the value functions but also on the future levels of emissions. Therefore it may happen that the threat of becoming an outsider to the smaller stable coalition is not sufficiently strong to deter deviations, even if the discount factor is close to one. Based on this we introduce the concept of *dynamic farsighted stability*. A coalition satisfies this type of stability if deviations are deterred, assuming that the deviator becomes an outsider to the smaller stable coalition that results *in the next period*, for a *different* level of emissions. The idea is similar to static farsighted stability. The set of stable coalitions is constructed from below. The size 1 coalition or the feedback Nash equilibrium between individual countries belongs to this set. By subsequently checking  $m = 2$ ,  $m = 3$  (if necessary), etcetera, the next element in the set can be found, in case it exists. If a next element is found, the procedure is continued until the full set of stable coalitions is characterized. This will become clear in the example that follows, but first we have to characterize the value function of the country that deviates from a size  $m$  coalition (which is the right hand side of inequality (26)).

Suppose therefore that a member of the coalition of size  $m$  considers deviating but the other countries in the coalition cannot detect this deviation and react to it before the next point in time. This implies that the deviating country can enjoy free-rider benefits

for one period of time but also has to face the consequences afterwards. The deviating country becomes an outsider to a smaller coalition of size  $m^+$  at the next point in time. The numbers  $m$  and  $m^+$  still have to be determined. Denote the parameter of the value function of an outsider to such a size  $m^+$  coalition by  $k^{o+}$ . It follows that the optimal behaviour of a deviating country is described by the dynamic programming equation

$$(28) \quad \frac{1}{2}k^d e^2 = \min \left[ \frac{1}{2}a_d^2 + \frac{1}{2}pe^2 + \frac{1}{2}\delta k^{o+} ((1+b)e - \sum_{k=1}^n a_k)^2 \right],$$

where  $a_d$  denotes the optimal abatement level of the deviating country and where  $V^d(e) = \frac{1}{2}k^d e^2$  denotes the value function of this country. Note that the last term in the dynamic programming equation (28) shows the different aspects that were discussed earlier. It indicates the discounted costs of becoming an outsider to a smaller coalition but for a lower level of emissions. The first-order condition becomes

$$(29) \quad a_d - \delta k^{o+} ((1+b)e - a_d - \sum_{k \neq d} a_k) = 0.$$

Because (in this time period) the other countries are assumed to stick to their strategy as a member of the coalition of size  $m$  or as an outsider to that coalition (equations (17) and (18), respectively), this leads to

$$(30) \quad a_d = \frac{\delta k^{o+}}{1 + \delta k^{o+}} \frac{1 + \delta m k^c}{1 + \delta(m^2 k^c + (n-m)k^o)} (1+b)e,$$

$$(31) \quad (1+b)e - \sum_{k=1}^n a_k = \frac{1}{1 + \delta k^{o+}} \frac{1 + \delta m k^c}{1 + \delta(m^2 k^c + (n-m)k^o)} (1+b)e.$$

It follows that the parameter of the value function of the deviating country is given by

$$(32) \quad k^d = p + \frac{\delta k^{o+}}{1 + \delta k^{o+}} \frac{(1 + \delta m k^c)^2}{(1 + \delta(m^2 k^c + (n-m)k^o))^2} (1+b)^2.$$

Deviations are deterred if  $k^c < k^d$ , where  $k^d$  is given by equation (32) and  $k^c$  (together with  $k^o$ ) is given by the set of equations (19) – (20).



In section 3 it was shown that for a static version of the game, a set of coalitions exists that satisfy static farsighted stability, with the largest coalition close to the grand coalition; in this section it was shown that this set of coalitions is also stable if the static game is repeated and if the discount factor is high enough. The question is whether a set of coalitions (larger than  $I$ ) exists that satisfy dynamic farsighted stability. In the sequel, the discount factor  $\delta$  is first taken equal to 1. Solutions to the dynamic programming equations still exist because of the special structure of the problem. At the end it is shown what the implications are of lowering the discount factor  $\delta$ .

The set of equations (19) - (20) cannot be solved analytically and therefore we have to resort to a numerical analysis. Values have to be given to three parameters: the factor  $b$ , indicating the business-as-usual growth in emissions, the number of countries  $n$ , and finally  $p$ , denoting the relative weight between damage and abatement costs. The value of  $b$  affects the abatement levels but it does not affect the qualitative results of the analysis. Therefore we only present the case  $b = 0$  here. The number of countries  $n$  is not essential for the qualitative results either. We have chosen  $n = 8$ , because then the full spectrum of results can be presented. This number is also not unrealistic, since international negotiations usually take place between blocks of countries. The value of  $p$ , however, is essential for the results. It will be shown that only for very low values of  $p$  a set of coalitions (larger than  $I$ ) exists that satisfy dynamic farsighted stability.

[Insert Tables 1 – 4]

Tables 1 – 4 show deviation costs associated with coalitions of size 1 through 8, with 8 representing the grand coalition, for four values of  $p$ . The first two rows in each table represent the solutions to the set of equations (19) – (20) for each value of  $m$ . For  $m = 1$ ,  $k^c$  and  $k^o$  are equal because this represents the feedback Nash equilibrium between individual countries. For  $m = 8$ , there are no outsiders. Note that a static analysis with respect to the value functions  $V^c$  and  $V^o$  gives the usual picture. Myopic stability gives a stable coalition of size 1 in all these cases, since  $k^c(m) > k^o(m - 1)$  for all  $m > 1$  (compare with equation (11) and the text thereafter). Static farsighted stability, however, always gives a set of stable coalitions larger than 1. In table 1 this set consists of size 1 and size 6 (since  $k^c(6) < k^o(1)$ ). In table 2 this set consists of size 1, size 3 (since  $k^c(3) < k^o(1)$ ) and size 7 (since  $k^c(7) < k^o(3)$ ). In tables 3 and 4 this set consists of size 1, size 3 (since  $k^c(3) < k^o(1)$ ) and size 6 (since  $k^c(6) < k^o(3)$ ). Note that the stability properties get better when the parameter  $p$  gets smaller. The stable set is larger in tables 2 – 4 than in table 1. Although the countries can coordinate on a larger coalition (size 7) in table 2 than in tables 3 – 4 (size 6), this is due to the fixed number of countries  $n$ : for a sufficiently large  $n$ , the stable set will prove to be larger in tables 3 – 4 than in table 2. The general conclusion is that in all these cases static farsighted stability allows the countries to coordinate on a large stable coalition.

In order to be able to investigate dynamic farsighted stability, we have to calculate  $k^d$  for the different coalition sizes (with equation (32)) and at the same time construct the stable set. The size 1 coalition or the feedback Nash equilibrium between individual countries belongs to this set. Therefore, in the first step of the algorithm  $m^+ = 1$  and  $k^{o+} = k^o(1)$ . Then  $k^d(2)$  can be calculated and compared to  $k^c(2)$ . If  $k^d(2) > k^c(2)$ , it follows that deviations are deterred and the coalition of size 2 is stable. If not,  $k^d(3)$

can be calculated and compared to  $k^c(3)$ , etcetera. Two things can happen: either an element  $m^*$  of the stable set is found or  $k^d(m) < k^c(m)$  for all  $m > 1$ . In the first case  $m^+$  is changed to  $m^*$  and  $k^{o+}$  is changed to  $k^o(m^*)$  and the procedure is continued for larger  $m$ . In the second case, the stable set consists of only the size 1 coalition. This last situation occurs in both tables 1 and 2. However, in table 3 first the coalition of size 3 is found to satisfy dynamic farsighted stability, since  $k^d(3) > k^c(3)$ , and then also the coalition of size 7, since  $k^d(7) > k^c(7)$ . In table 4 the stable set consists of size 3 and size 6, since  $k^d(3) > k^c(3)$  and  $k^d(6) > k^c(6)$ . In contrast to the conclusion above under static farsighted stability, now it is found that countries can only coordinate on a large stable coalition, if the parameter  $p$  is very small (tables 3 and 4).

It is time to take stock of what we have found. Initially the usual picture reappears. The concept of myopic stability gives very small stable coalitions, but the concept of static farsighted stability gives both small and large stable coalitions and allows the countries to coordinate on the largest stable coalition. This conclusion partly changes, however, if it is assumed that detection of deviations takes some time. The reason is that in that time abatement occurs and therefore the level of emissions decreases prior to detection. It follows that costs will have decreased, and therefore the threat of triggering a smaller coalition may not be sufficient to deter defections anymore. It can be seen in tables 1 – 4 that deviations are only deterred when the parameter  $p$  is very small (as in tables 3 and 4). It is interesting to see that the value of the parameter  $p$  again determines whether a coalition is stable or not, as in the case of myopic stability for the simple static model in section 3. While the intuition in the simple model was based on the steepness of the best-reply function, here there is an additional aspect. The parameter  $p$  denotes the relative weight of the costs attached to the emission

levels. If  $p$  is small, the level of emissions is slowly adjusted and therefore the game changes slowly over time. If  $p$  is large, however, the level of emissions and therefore costs decrease substantially in each time period. Apparently, in the last situation the trigger mechanism does not sustain large coalitions anymore. Note that in the case of myopic stability for the simple static model in section 3, the distinction that  $p$  made was only between a size 2 and a size 1 coalition; with dynamic farsighted stability in a full dynamic model, the distinction is between a large coalition and a size 1 coalition.

It is interesting to see what the implication is of lowering the discount factor  $\delta$ . It has an effect on all the costs, of course, but it can also have an effect on the set of stable coalitions. If the parameter  $p$  is relatively large, as in the context of tables 1 and 2, a reduction of the discount factor  $\delta$  from 1.0 to 0.9 does not change the set of coalitions that satisfy dynamic farsighted stability: it still only consists of the size 1 coalition. If, however,  $p$  is sufficiently small, as in the context of tables 3 and 4, the set of stable coalitions consists of size 1, size 2, size 4 and size 7, following a reduction of  $\delta$  to 0.9. Thus, if a set of stable coalitions larger than 1 exists, this set contains more elements for  $\delta = 0.9$  than for  $\delta = 1$ .

In some sense, the circle is closed. The static model in section 3 led to the following conclusion, formulated by Finus [11]: the steeper the best-reply function, the lower the number of participants. Since the parameter  $p$  is an indicator for the steepness of the best-reply function, the conclusion can be reformulated as follows: the higher  $p$ , the lower the size of the coalition. Under static farsightedness, the dependence on  $p$  had disappeared. This paper shows, however, that the conclusion is re-established by

integrating the behavioural reaction patterns and the emissions adjustment processes in one dynamic context: only if  $p$  is very small, large coalitions can be sustained.

It is encouraging that this conclusion gets wide support with different types of models. On the other hand, the conclusion does not seem to reflect the practice of International Environmental Agreements. It is generally argued ([3,11]) that the Montreal Protocol can be characterized by a high value of  $p$ , since abatement costs are relatively small as compared to the costs of holes in the ozone layer. The Kyoto Protocol, on the other hand, can be characterized by a medium value of  $p$ , since both abatement costs and costs of climate change are high. Therefore, it should be easier to sustain the Kyoto Protocol, but the opposite has occurred. It can be argued that  $p$  for the Kyoto Protocol is not small enough to sustain a large coalition, but this does not explain why the Montreal Protocol has been a success. The reason is probably that due to technological development, the issue disappeared ([2]): it proved not to be difficult to do without CFC's. The Agreement was interesting for many reasons but not so much because it solved a problem. Theory may also give an explanation. Barrett [2] shows for the static model that the result turns around if for the game between the coalition and the outsiders, the Stackelberg concept is used instead of the Nash equilibrium. Finus [11] also concludes the Stackelberg assumption yields a larger number of participants, the steeper is the best-reply function; this result is in line with the Montreal Protocol. Finus [11] is critical of the Stackelberg concept in this context, but it may anyhow be interesting to investigate whether the result for the dynamic model also turns around, if the Stackelberg concept is used. This question is left for further research. At this point it is concluded that theory predicts that it will be hard to sustain a large coalition, unless damage costs are very small as compared to abatement costs.

This is bad news for International Environmental Agreements, but it only stresses the importance to design agreements in a way that may overcome these mechanisms. This is also left for further research.

## **5. Conclusion**

Stability of international environmental agreements is a heavily debated issue. Some are optimistic and expect that the grand coalition can be sustained. Their arguments are based on the core concept in cooperative game theory or on trigger mechanisms in repeated games. Others are pessimistic and expect that only very small coalitions can be sustained. They base themselves on the concept of internal and external stability, which is called myopic stability in this paper. The idea of farsightedness can be used to reconcile these approaches: both large and small stable coalitions can occur, and countries may be able to coordinate on a large stable coalition. This is good news for the optimists.

This paper first clarifies these issues and then takes the analysis one step further. The main issue is compliance or deterrence of deviations. This paper shows that if it takes time to detect deviations, trigger mechanisms may not work. The reason is simply that the game changes over time. This leads to the result that large coalitions can only be sustained, if damage costs are very small as compared to abatement costs. In static models under myopic stability a similar result appeared, but now the result is stronger and it occurs in a much richer dynamic framework. The result implies that in the most important situations, in which environmental damage costs are relatively substantial, large coalitions will not easily arise. Unfortunately, the pessimistic view still prevails.

The challenge remains to find ways in which stability properties of large coalitions can be improved. This is left for further research.

## References

- [1] C. d'Aspremont, J. Jacquemin, J. Gabszewicz, J. Weymark, On the stability of collusive price leadership, *Canadian Journal of Economics* 16 (1983) 17-25.
- [2] S. Barrett, Self-enforcing international environmental agreements, *Oxford Economic Papers* 46 (1994) 878-894.
- [3] S. Barrett, Environment and Statecraft, The Strategy of Environmental Treaty-making, Oxford University Press, Oxford, 2003.
- [4] F. Bloch, Non-cooperative models of coalition formation in games with spillovers, in: C. Carraro, D. Siniscalco (Ed.), *New Directions in the Economic Theory of the Environment*, Cambridge University Press, Cambridge, 1997, pp. 311-352.
- [5] C. Carraro, D. Siniscalco, Strategies for the international protection of the environment, *Journal of Public Economics* 52 (1993) 309-328.
- [6] P. Chander, H. Tulkens, A core-theoretic solution for the design of cooperative agreements on transfrontier pollution, *International Tax and Public Finance* 2 (1995) 279-293.

- [7] M. Chwe, Farsighted coalitional stability, *Journal of Economic Theory* 63 (1994) 299-325.
- [8] E. Diamantoudi, E. Sartzetakis, International environmental agreements – the role of foresight, Working paper, 2002.
- [9] E. Dockner, S. Jørgensen, N. Long, G. Sorger, *Differential Games in Economics and Management Science*, Cambridge University Press, Cambridge, U.K., 2000.
- [10] J. Eyckmans, On the farsighted stability of international climate agreements, Working paper, 2003.
- [11] M. Finus, Stability and design of international environmental agreements: the case of transboundary pollution, in: H. Folmer, T. Tietenberg (Ed.), *The International Yearbook of Environmental and Resource Economics 2003/2004*, Edward Elgar, Cheltenham, 2003, pp. 82-158.
- [12] J. Friedman, *Game Theory with Applications to Economics*, Oxford Economic Press, New York, 1986.
- [13] M. Hoel, International environmental conventions: the case of uniform reductions of emissions, *Environmental & Resource Economics* 2 (1992) 141-159.
- [14] L. Karp, J. Zhao, A proposal to reform the Kyoto Protocol: the role of escape clauses and foresight, Working paper, 2007.



- [15] F. van der Ploeg, A. de Zeeuw, International aspects of pollution control, *Environmental & Resource Economics* 2 (1992) 117–139.
- [16] D. Ray, R. Vohra, Coalitional power and public goods, *Journal of Political Economy* 109 (2001) 1355-1384.
- [17] S.-S. Yi, Stable coalition structures with externalities, *Games and Economic Behaviour* 20 (1997) 201-237.

**Table 1** Insider/outsider/deviator costs, coalition size  $m$ ,  $p = 1$  ( $b = 0$ ,  $n = 8$ )

	m=1***	m=2	m=3	m=4	m=5	m=6*	m=7	m=8
$k^c$	1.0245	1.0425	1.0456	1.0391	1.0311	1.0244	1.0192	1.0154
$k^o$	1.0245	1.0161	1.0085	1.0043	1.0023	1.0013	1.0007	
$k^d$		1.0379	1.0363	1.0287	1.0216	1.0162	1.0124	1.0097

**Table 2** Insider/outsider/deviator costs ( $10^{-2}$ ), coalition size  $m$ ,  $p = 10^{-2}$  ( $b = 0$ ,  $n = 8$ )

	m=1***	m=2	m=3*	m=4	m=5	m=6	m=7*	m=8
$k^c$	3.0169	3.1472	3.0143	2.7150	2.4231	2.1841	1.9953	1.8463
$k^o$	3.0169	2.6455	2.1598	1.8128	1.5913	1.4449	1.3434	
$k^d$		3.0050	2.8304	2.5850	2.3464	2.1370	1.9594	1.8105

**Table 3** Insider/outsider/deviator costs ( $10^{-3}$ ), coalition size  $m$ ,  $p = 10^{-4}$  ( $b = 0$ ,  $n = 8$ )

	m=1***	m=2	m=3***	m=4	m=5	m=6*	m=7**	m=8
$k^c$	2.6215	2.7386	2.5829	2.2505	1.9307	1.6702	1.4644	1.3010
$k^o$	2.6215	2.2307	1.6960	1.3082	1.0582	0.8896	0.7692	
$k^d$		2.6192	2.5945	1.6899	1.6612	1.6317	1.6022	0.7687

**Table 4** Insider/outsider/deviator costs ( $10^{-4}$ ), coalition size  $m$ ,  $p = 10^{-5}$  ( $b = 0$ ,  $n = 8$ )

	m=1***	m=2	m=3***	m=4	m=5	m=6***	m=7	m=8
$k^c$	8.2042	8.5716	8.0749	7.0175	6.0010	5.1741	4.5212	4.0032
$k^o$	8.2042	6.9650	5.2650	4.0319	3.2379	2.7027	2.3207	
$k^d$		8.2019	8.1765	5.2585	5.2287	5.1977	2.7013	2.6851

\*: static farsighted stable; \*\*: dynamically farsighted stable

---

<sup>1</sup> The author is grateful to Gérard Gaudet, Reyer Gerlagh, Michael Hoel, Eric Maskin, Alistair Ulph, three anonymous referees, the editor, and seminar participants in Victoria, Paris, New Delhi, Antwerp and Rethymnon for comments on earlier versions of this paper; also the hospitality of CAS in Oslo is gratefully acknowledged.

<sup>2</sup> The idea is that output or income grows and related to that, emissions grow; in order to lower emissions, decoupling is needed which means that abatement has to realize a larger percentage decrease in emissions than the percentage increase in emissions that results in the absence of environmental concern.

<sup>3</sup> Thanks are due to one of the referees for suggesting these names.

<sup>4</sup> This assumption is not self-evident as can be seen in Karp and Zhao (2007).